# Towards Automatic Abdominal Multi-Organ Segmentation in Dual Energy CT using Cascaded 3D Fully Convolutional Network

Shuqing Chen, Holger Roth, Sabrina Dorn, Matthias May, Alexander Cavallaro, Michael Lell, Marc Kachelrieß, Hirohisa Oda, Kensaku Mori and Andreas Maier

***Abstract*—Automatic multi-organ segmentation of the dual energy computed tomography (DECT) data can be beneficial for biomedical research and clinical applications. Recent advances in deep learning showed the feasibility to use 3D fully convolutional networks (FCN) for voxel-wise dense predictions in single energy computed tomography (SECT). In this paper, we propose a 3D FCN based method for automatic multi-organ segmentation in DECT. The work was based on a cascaded FCN for the major organs trained on a large set of SECT data. We preprocessed the DECT data by using linear weighting and fine-tuned the FCN for the DECT data. The method was evaluated using 42 torso DECT data acquired with a clinical dual-source CT system. Four abdominal organs (liver, spleen, left and right kidneys) were evaluated with cross-validation strategy. Effect of the weight on the accuracy was researched. In all the tests, we achieved an average Dice coefficient of 93% for the liver, 92% for the spleen, 91% for the right kidney and 89% for the left kidney, respectively. The results show that our method is feasible and promising.**

## I. Introduction

The Hounsfield unit (HU) scale value depends on the inherent tissue properties, the x-ray spectrum for scanning and the administered contrast media [1]. In a SECT image, materials having different elemental compositions can be represented by identical HU values [2]. Therefore, SECT has challenges such as limited material-specific information and beam hardening as well as tissue characterization [1]. DECT has been investigated to solve the challenges of SECT. In DECT, two energy-specific image data sets are acquired at two different X-ray spectra, which are produced by different energies, simultaneously. The multi-organ segmentation in DECT can be beneficial for biomedical research and clinical applications, such as material decomposition [3], organ-specific context-sensitive enhanced reconstruction and display [4], [5], and computation of bone mineral density [6]. We are aiming at exploiting the prior anatomical information that is gained through the multi-organ segmentation to provide an improved context-sensitive DECT imaging [4], [5]. The novel technique offers the possibility to present evermore complex information to the radiologists simultaneously and bears the potential to improve the clinical routine in CT diagnosis.

Automatic multi-organ segmentation on DECT images is a challenging task due to the inter-subject variance of human abdomen, the complex 3D intra-subject variance among organs, soft anatomy deformation, as well as different HU values for the same organ by different spectra. Recent researches show the power of deep learning in medical image segmentation [7]. To solve the DECT segmentation problem, we use the successful experience from multi-organ segmentation in volumetric SECT images using deep learning [8], [9]. The proposed method is based on a cascaded 3D FCN, a two-stage, coarse-to-fine approach [8]. The first stage is used to predict the region of the interest (ROI) of the target organs, while the second stage is learned to predict the final segmentation. No organ-specific or energy-specific prior knowledge is required in the proposed method. The cross-validation results showed that the proposed method is promising to solve multi-organ segmentation problem for DECT. To the best of our knowledge, this is the first study about multi-organ segmentation in DECT images based on 3D FCNs.

## II. Materials and Methods

### A. Network Architecture for DECT Prediction

As described by Krauss et al. [10], a mixed image display is employed in clinical practice for the diagnose using DECT. The mixed image is calculated by linear weighting of the images values of the two spectra:

$$I_{\text{mix}} = \alpha \cdot I_{\text{low}} + (1 - \alpha) \cdot I_{\text{high}} \tag{1}$$

where $\alpha$ is the weight of the dual energy composition, $I_{\text{mix}}$ denotes the mixed image. $I_{\text{low}}$ and $I_{\text{high}}$ are the images at low and high kV, respectively.

We preprocessed the DECT images following Eq. 1 straightforwardly. Figure 1 illustrates the network architecture of the proposed method for the DECT multi-organ segmentation. To prepare network training, labeled segmentation is generated manually by experts for each training data. In the training phase, first of all, mixed image is calculated by combining the images at the low energy level and the high energy level using Eq. 1. Then, a binary mask is generated by thresholding the skin contour of the mixed image. Subsequently, the mixed

S. Chen is with Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany (e-mail: shuqing.chen@fau.de).

H. Roth, H. Oda, and K. Mori are with Nagoya University, Nagoya , Japan.

S. Dorn, and M. Kachelrieß are with German Cancer Research Center (DKFZ), Heidelberg, Germany, and also with Ruprecht-Karls-University Heidelberg, Heidelberg, Germany.

M. May, and A. Cavallaro are with Department of Radiology, University Hospital Erlangen, Erlangen, Germany.

M. Lell is with University Hospital Nürnberg, Paracelsus Medical University, Nürnberg, Germany.

A. Maier is with Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany.

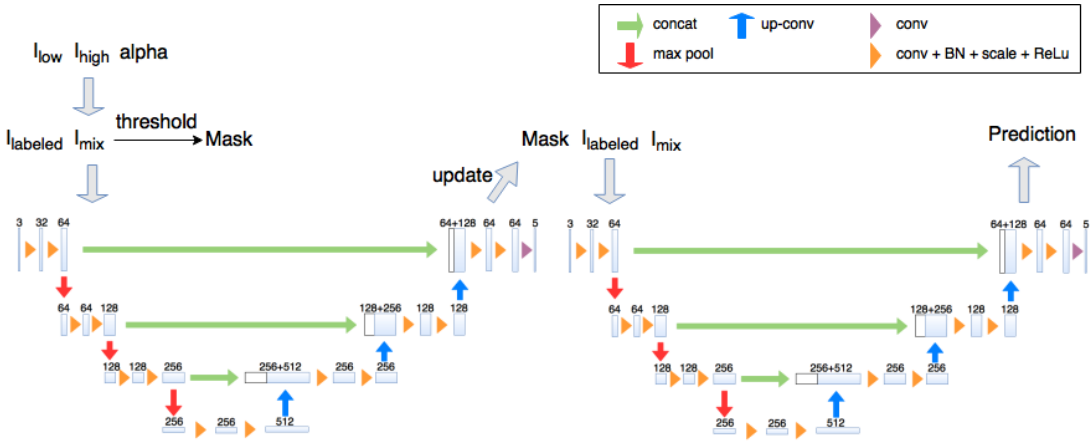


Fig. 1: Cascaded network architecture for DECT multi-organ segmentation

image, the binary mask and the labeled image are given into the network as multi-channel inputs. The network consists of two stages. The first stage is applied to generate the region of the interest (ROI) in order to reduce the search space for the second stage. The prediction result of the first stage is taken as the mask for the second stage. Each stage is based on a standard 3D U-Net [11], which is a fully convolutional network including an analysis and a synthesis path. We used the open-source implementation of two stages cascaded network [8] developed by Roth et al. based on the 3D U-Net [11] and the Caffe deep learning library [12]. The cascaded network was trained by Roth et al. [8] on a large set of SECT images including some of the major organ labels. Our model was trained by fine-tuning the pre-trained network with the mixed DECT images using the pre-trained weights as initialization. The difference between the network output and the ground truth labels are compared using softmax with weight voxel-wise cross-entropy loss [8], [11].

### B. Experimental Setup

The proposed method was evaluated with 42 clinical torso DECT images scanned by the department of radiology, university hospital Erlangen. All of the images were taken from male and female adult patients who had different clinically oriented indication justified by the radiologist. Ultravist 370 was given as contrast agent with body weight adapted volumes. The images were acquired at different X-ray tube voltage setting of 70 kV (560 mAs) and Sn 150 kV (140 mAs, with Sn filter) using a Siemens SOMATOM Force CT system with Stellar, an energy integrating detector. The traning volumes contains 992-1290 slices with slice size 512x512 pixels. The voxel dimensions are [0.6895-0.959, 0.6895-0.959, 0.6] mm. Four abdominal organs were tested, including liver, spleen, right and left kidneys. Ground truth was generated by experts manually.

To avoid the bias of the data selection and to keep the dataset distribution similar, a manifold learning-based technique [13] was applied to split the data into training dataset, validation dataset, and test dataset. First, the images were resized to the same image spacing (e.g.[3mm 3mm 5mm]). Then, the distribution of the images was calculated and plotted by using locally linear embedding (LLE) [14]. Subsequently, the images were clustered into 3 classes using k-means. Finally, training data, validation data, and test data were selected randomly from these classes with the ratio 5:1:1, i.e. in each test we used 2 images from each class (6 in total) for validation, 2 images from each class for test (6 in total), and the remaining 30 images for training.

| | | Liver | Spleen | r.Kidney | l.Kidney |
|---|---|---|---|---|---|
| DECT | Avg. | 0.92 | 0.84 | 0.88 | 0.87 |
| | SD | 0.02 | 0.08 | 0.03 | 0.03 |
| | Min. | 0.84 | 0.62 | 0.80 | 0.78 |
| | Max. | 0.94 | 0.95 | 0.94 | 0.93 |

TABLE I: Dice coefficients of cross-validation with $\alpha_{\text{training}}$=0.6 and $\alpha_{\text{test}}$ =0.6. SD is abbreviated for standard deviation.

## III. RESULTS

### A. Performance Estimation with Cross-Validation

NVIDIA GeForce GTX 1080 Ti with 11 GB memory was used for all of the experiments. The similarity between the segmentation result and the ground truth was measured with Dice metric by using the tool provided by VISCERAL [15]. First, the performance of the proposed method was estimated by 8-folds cross-validation, using 0.6 as $\alpha_{\text{training}}$ as well as $\alpha_{\text{test}}$. Fig. 2 shows one segmentation results in 3D. TableI summarizes the Dice coefficients of the segmentation results and compares DECT results with the SECT results. The proposed method under the above weight condition yielded an average Dice coefficient of 92% for the liver, 84% for the spleen, 88% for the right kidney and 87% for the left kidney, respectively. Fig. 3 plots the distributions of the Dice coefficients for different test scenarios and showed the high robustness of the proposed method. Though the Dice coefficients under above mentioned weight condition are less than SECT results in [9], we performed a second test which is focused on the weight alpha both for training and for test phase.

### B. Study on the Weight $\alpha$

We are aiming at exploiting the spectral information in the DECT data. Since the $\alpha$ mixing results basically in pseudo monochromatic images comparable to single energy scans, the influence of the weight $\alpha$ on the accuracy was further

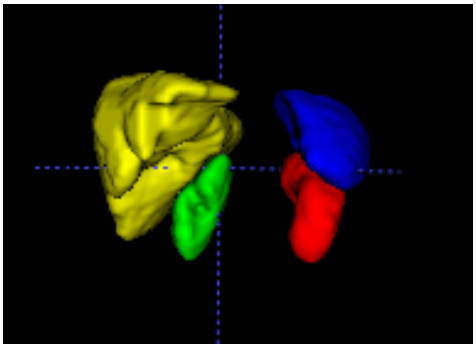

Fig. 2: 3D rendering of one DECT segmentation with yellow for liver, blue for spleen, green for right kidney and red for

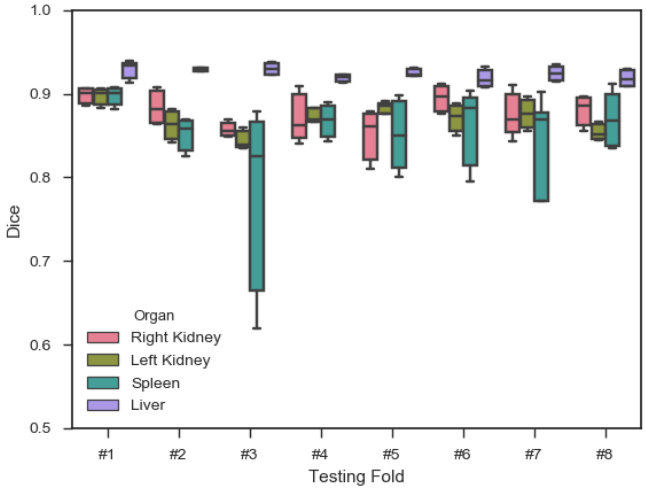


Fig. 3: Dice coefficients of the target organs with $\alpha_{\text{training}} = 0.6$ and $\alpha_{\text{test}} = 0.6$ for 8 different testing folds

researched. 0, 0.3, 0.6, 0.9 and 1 were chosen as $\alpha_{\text{training}}$ and $\alpha_{\text{test}}$ in this study. Table II lists the average Dice coefficient. For all of the cases, the liver had the highest accuracy (92%-93%). The segmentation of the right kidney was usually more accurate than the left kidney. The best Dice values per organ per training set are highlighted in Table II. The test with $\alpha_{\text{training}}$=0.9 and $\alpha_{\text{test}}$=0.9 obtained the highest accuracy for liver and right kidney. The test with weight combination 0.9-1 showed the best segmentation for spleen, the combination with 0.9-0.3 had the finest result for left kidney. High $\alpha_{\text{test}}$ generated better segmentation for liver and spleen. For most organs, the best Dice values of DECT are higher than the SECT results given in [9].

## IV. DISCUSSION AND CONCLUSION

We proposed a deep learning based method for automatic abdominal multi-organ segmentation in DECT. The evaluation results show the feasibility of the proposed method. Compared to the results of the SECT images reported by Roth et al. [9], our method is promising and robust (see Table II). For most organs, the segmentation of our method is more accurate than the SECT [9] when an optimal fusion weight is selected. The results illustrate that the image fusion affects the segmentation of DECT. In the cross validation, the third testing fold had a large deviation. The reason could be that our image data were

| $\alpha_{\text{training}}$-$\alpha_{\text{test}}$ | Liver | Spleen | r.Kidney | l.Kidney |
|---|---|---|---|---|
| 0-0 | 0.908 | *0.878* | 0.840 | *0.852* |
| 0-0.3 | 0.915 | 0.876 | 0.860 | 0.841 |
| 0-0.6 | 0.919 | 0.875 | *0.865* | 0.839 |
| 0-0.9 | 0.922 | 0.876 | 0.864 | 0.837 |
| 0-1 | *0.923* | 0.876 | 0.861 | 0.835 |
| 0.3-0 | 0.876 | 0.885 | 0.845 | 0.835 |
| 0.3-0.3 | 0.924 | 0.899 | *0.900* | *0.891* |
| 0.3-0.6 | 0.925 | *0.902* | 0.891 | 0.881 |
| 0.3-0.9 | *0.926* | 0.901 | 0.877 | 0.859 |
| 0.3-1 | 0.921 | 0.900 | 0.877 | 0.854 |
| 0.6-0 | 0.865 | 0.857 | 0.786 | 0.796 |
| 0.6-0.3 | 0.909 | 0.897 | 0.844 | 0.885 |
| 0.6-0.6 | *0.922* | 0.904 | *0.895* | *0.887* |
| 0.6-0.9 | 0.912 | 0.906 | *0.895* | 0.873 |
| 0.6-1 | 0.919 | *0.908* | 0.843 | 0.866 |
| 0.9-0 | 0.881 | 0.848 | 0.745 | 0.764 |
| 0.9-0.3 | 0.930 | 0.901 | 0.898 | **0.892** |
| 0.9-0.6 | 0.932 | 0.908 | 0.904 | 0.873 |
| 0.9-0.9 | **0.933** | 0.915 | **0.906** | 0.862 |
| 0.9-1 | 0.930 | **0.917** | 0.905 | 0.872 |
| 1-0 | 0.907 | 0.822 | 0.784 | 0.812 |
| 1-0.3 | 0.912 | 0.886 | 0.869 | 0.889 |
| 1-0.6 | 0.915 | 0.895 | *0.879* | 0.886 |
| 1-0.9 | 0.917 | 0.901 | *0.879* | *0.891* |
| 1-1 | *0.918* | *0.902* | 0.877 | *0.891* |
| **SECT** [9] | 0.95 | 0.90 | 0.90 | 0.88 |

TABLE II: Dice coefficients of different alpha for testing fold 1. Bold denotes the best organ results of DECT. Italic underline denotes the best results in the group with the same training weight. Notice that the DECT and SECT approaches used different data set.

taken from patients with different disease (liver tumor, spleen tumor, etc.). The disease type is not considered by the data selection. Training and test with inconsistent symptoms could have an impact on the accuracy.

The study on the weight can be divided into three groups with different $\alpha_{\text{training}}$. $\alpha$=0.9 is close to the low energy images which have on average the best soft-tissue contrast, $\alpha_{\text{training}}$=0.9 worked thus better in general. The intra-group comparison showed that the cases with identical training and test conditions had a higher probability to get the best segmentation result. This is expected because the mixed images generated by the matched training and test conditions may have the highest similarity. Furthermore, the comparison of the case 0.3-0.9 (low-contrast model for high-contrast image) with the case 0.9-0.3 (high-contrast model for low-contrast image) showed that using a model trained on high-contrast images for segmenting low-contrast test images works better. In addition, liver is well segmented in middle to high $\alpha$ ranges. Spleen is segmented best at $\alpha$=0.6. Kidneys work best in matched training and test conditions. This suggests that there is an optimal $\alpha$ for each organ for image segmentation.

The weight $\alpha$ for the mixed image calculation is currently a user-defined parameter in the preprocessing in our approach. The fact suggests that the alpha shall be regarded as organ specific parameter in the network and optimized in the training phase. It can be used to augment the data for the training in future. Also, the network could be modified with two image inputs. Furthermore, more organs and more scans from different patients could be used.

**ACKNOWLEDGMENTS** This work was supported by the

German Research Foundation (DFG) through research grant No. KA 1678/20, LE 2763/2-1 and MA 4898/5-1.

## REFERENCES


[1] Dushyant Sahani, "Dual-energy CT: The technological approaches," Society of Computed Body Tomography and Magnetic Resonance (SCBT-MR), 2012.

[2] Cynthia H. McCollough, Shuai Leng, Lifeng Yu, and Joel G. Fletcher, "Dual- and multi-energy CT: Principles, technical approaches, and clinical applications," *Radiology*, vol. 276, no. 3, pp. 637–653, 2015.

[3] Stefan Kuchenbecker, Sebastian Faby, David Simons, Michael Knaup, Heinz-Peter Schlemmer, Michael M. Lell, and Marc Kachelriess, "Segmentation-assisted material decomposition in dual energy computed tomography (DECT)," in *Radiological Society of North America (RSNA)*, 2015.

[4] Sabrina Dorn, Shuqing Chen, Francesco Pisana, Joscha Maier, Michael Knaup, Stefan Sawall, Andreas Maier, Michael Lell, and Marc Kachelrieß, "Organ-specific context-sensitive single and dual energy CT (DECT) image reconstruction, display and analysis," in *103rd Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA)*, 2017.

[5] Sabrina Dorn, Shuqing Chen, Stefan Sawall, David Simons, Matthias May, Joscha Maier, Michael Knaup, Heinz-Peter Schlemmer, Andreas Maier, Michael Lell, and Marc Kachelrieß, "Organ-specific context-sensitive CT image reconstruction and display," in *Medical Imaging Proc. SPIE*, 2018.

[6] S Wesarg, M Kirschner, M Becker, M Erdt, K Kafchitsas, MF Khan, et al., "Dual-energy CT-based assessment of the trabecular bone in vertebrae," *Methods of information in medicine*, vol. 51, no. 5, pp. 398, 2012.

[7] Marc Aubreville, Miguel Goncalves, Christian Knipfer, Nicolai Oetter, Tobias Würfl, Helmut Neumann, Florian Stelzle, Christopher Bohr, and Andreas K. Maier, "Patch-based carcinoma detection on confocal laser endomicroscopy images - A cross-site robustness assessment," *CoRR*, vol. abs/1707.08149, 2017.

[8] Holger R. Roth, Hirohisa Oda, Yuichiro Hayashi, Masahiro Oda, Natsuki Shimizu, Michitaka Fujiwara, Kazunari Misawa, and Kensaku Mori, "Hierarchical 3D fully convolutional networks for multi-organ segmentation," in *arXiv preprint arXiv:1704.06382*.

[9] Holger Roth, Ying Yang, Masahiro Oda, Hirohisa Oda, Yuichiro Hayashi, Natsuki Shimizu, Takayuki Kitasaka, Michitaka Fujiwara, Kazunari Misawa, and Kensaku Mori, "Torso organ segmentation in CT using fine-tuned 3D fully convolutional networks," in *36 (JAMIT)*, 2017.

[10] Bernhard Krauss, Bernhard Schmidt, and Thomas G. Flohr, *Dual Energy CT in Clinical Practice*, chapter Dual Source CT, Springer Berlin Heidelberg, 2011.

[11] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2016.

[12] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[13] Shuqing Chen, Sabrina Dorn, Michael Lell, Marc Kachelrieß, and Andreas Maier, "Manifold learning-based data sampling for model training," in *Informatik-Aktuell*, Bildverarbeitung für die Medizin: Algorithmen-Systeme-Anwendungen, Ed., Germany, 2018, pp. 269–274.

[14] Sam T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[15] Abdel Aziz Taha and Allan Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, pp. 29, August 2015.